Schools can use data from the Healthy Youth Survey 2006 to:

- learn the **prevalence** of health-related behaviors among youth;
- identify **changes** over time; and
- **compare** groups, such as youth in a school to youth in the state as a whole.

However, the usefulness of the data depends on several factors, described below.

Why are these issues important to understand?

The Healthy Youth Survey 2006 for Grades 8, 10 and 12 comes in two forms. Every student will get one of the two forms in alternating order. Both forms have a core set of 35 questions. Thus, when all the students in a given grade take the survey, there will be a *census* for the core questions. For the remaining questions, there will be a 50% *sample*, because half the students will answer questions on one form and half will answer questions on the other. For Grade 6, there is only one form and so there will be a *census*.

Census or sample?

Census. When all of the students answer a question, it is called a census. A census describes the people who took the survey. For example, if you want to know how many students in Grade 8 smoke to decide whether to start a smoking cessation program, a census will give you the number who report smoking and nothing else may be required. However, there are additional considerations if you wish to generalize beyond the students who took the survey. For example, if you want to start a program for smokers in Grade 8 next year, you cannot assume the number of smokers will be the same as this year. You need confidence intervals to know how closely the students next year can be expected to resemble the students this year.

Sample. In a sample, only some of the students answer a question, but *the sample needs to represent all of the students.* Procedures to choose a random sample—such as giving every other student one of two forms—yield a sample that represents all the students. Procedures that are not random—such as giving the survey only to honors students or giving the survey on a day when band members are out of school—can result in a sample that is not representative. If the sample is chosen by a valid procedure and there are not large amounts of missing data, confidence intervals give an estimate of how accurately you can generalize from the sample to the larger population.

What are confidence intervals?

Confidence intervals give an estimate of how variable the results are. Specifically, the 95% confidence interval gives the range that contains the true value 95% of the time. In the smoking example above, the percent of students in Grade 8 who would report smoking if they were surveyed next year should be within the range of the 95% confidence intervals, if there have not been major changes in smoking rates.

The size of the confidence intervals in the Healthy Youth survey will be smaller when

- there are more students taking the survey and

- the prevalence rates are farther from a 50%/50% split.

For example, if 200 students take the survey and 30% report smoking (a core item), the 95% confidence interval will be around 24%–36% (that is, about ± [plus or minus] 6%). However, only half of the 200 students will answer the question on bicycle helmets, as this question is only on one of the forms. If 30% of the 100 students report wearing bicycle helmets, the 95% confidence interval will be around 21%–39% (that is, ± 9%).

Why conduct significance tests?

Sometimes, people look at **trends** over time or **compare two groups**. A statistical test can tell you whether the difference between groups or over time is greater than would be expected by chance. Commonly, if a difference as large as the one we see occurs only 5% of the time by chance, we say that the difference is statistically significant. As the number of students taking the survey gets smaller, you need larger differences to rule out chance. For example, using a common statistical test, a statistically significant difference can be found comparing 30% to about:

- 21% if there are 200 students per group;
- 17% if there are 100 students per group; and
- 7% if there are only 30 students per group.

If a difference is not statistically significant it may be due to chance, making it more difficult to interpret.

Another way to examine differences is to see whether the confidence intervals overlap. If they do not overlap, the difference is probably not due to chance. However, this is not an exact method, and there may be a statistically significant difference even when two confidence intervals overlap.

What if small numbers make my data difficult to interpret?

Combining your data with that from other schools in your district or county will increase the sample size and may help with data interpretation. Additionally, because data may be used at the district or county level, it is important that your school contribute to district or county information even if it is difficult to interpret data from your school alone. The considerations discussed above also apply to school districts and counties.

Summary

To help interpret data, reports of the Healthy Youth Survey 2006 will include 95% confidence intervals. As long as there is not a lot of missing data (such as non-participating classrooms) and there are at least 15 valid surveys per grade, the Healthy Youth Survey 2006 can provide **prevalence** estimates of behaviors among students at your school. Generally, larger schools will have smaller confidence intervals, indicating more precise estimates.

In looking at **trends** over time or **comparing two groups**, confidence intervals will give an idea of how much difference is needed to detect a difference, but more formal statistical tests may be needed. Smaller confidence intervals and more power to detect statistically significant differences can be obtained by combining your data with that from other schools.